

Technical University of Denmark



Systematic comparison of different techniques to measure hippocampal subfield volumes in ADNI2

Mueller, Susanne G.; Yushkevich, Paul A.; Das, Sandhitsu; Wang, Lei; Van Leemput, Koen; Iglesias, Juan Eugenio; Alpert, Kate; Mezher, Adam; Ng, Peter; Paz, Katrina; Weiner, Michael W.

Published in:
NeuroImage: Clinical

Link to article, DOI:
[10.1016/j.nicl.2017.12.036](https://doi.org/10.1016/j.nicl.2017.12.036)

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Mueller, S. G., Yushkevich, P. A., Das, S., Wang, L., Van Leemput, K., Iglesias, J. E., ... Weiner, M. W. (2018). Systematic comparison of different techniques to measure hippocampal subfield volumes in ADNI2. NeuroImage: Clinical, 17, 1006-1018. DOI: 10.1016/j.nicl.2017.12.036

DTU Library

Technical Information Center of Denmark

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Systematic comparison of different techniques to measure hippocampal subfield volumes in ADNI2

Susanne G. Mueller^{a,*}, Paul A. Yushkevich^b, Sandhitsu Das^b, Lei Wang^c, Koen Van Leemput^{d,e}, Juan Eugenio Iglesias^{d,f}, Kate Alpert^c, Adam Mezher^g, Peter Ng^g, Katrina Paz^g, Michael W. Weiner^a, for the Alzheimer's Disease Neuroimaging Initiative¹

^a Dept. of Radiology and Biomedical Imaging, University of California, San Francisco, CA, USA

^b Penn Image Computing and Science Laboratory, Dept. of Radiology, University of Pennsylvania, Philadelphia, PA, USA

^c Northwestern University, Feinberg School of Medicine, Chicago, IL, USA

^d Athinoula A. Martinos Center for Biomedical Imaging, Boston, MA, USA

^e Dept. of Applied Mathematics and Computer Science, Technical University of Denmark, Denmark

^f Translational Imaging Group, University College London, London, UK

^g Center for Imaging of Neurodegenerative Diseases (CIND), VAMC San Francisco, San Francisco, CA, USA

A B S T R A C T

Objective: Subfield-specific measurements provide superior information in the early stages of neurodegenerative diseases compared to global hippocampal measurements. The overall goal was to systematically compare the performance of five representative manual and automated T1 and T2 based subfield labeling techniques in a subset of the ADNI2 population.

Methods: The high resolution T2 weighted hippocampal images (T2-HighRes) and the corresponding T1 images from 106 ADNI2 subjects (41 controls, 57 MCI, 8 AD) were processed as follows. A. T1-based: 1. FreeSurfer + Large-Diffeomorphic-Metric-Mapping in combination with shape analysis. 2. FreeSurfer 5.1 subfields using in-vivo atlas. B. T2-HighRes: 1. Model-based subfield segmentation using ex-vivo atlas (FreeSurfer 6.0). 2. T2-based automated multi-atlas segmentation combined with similarity-weighted voting (ASHS). 3. Manual subfield parcellation. Multiple regression analyses were used to calculate effect sizes (ES) for group, amyloid positivity in controls, and associations with cognitive/memory performance for each approach.

Results: Subfield volumetry was better than whole hippocampal volumetry for the detection of the mild atrophy differences between controls and MCI (ES: 0.27 vs 0.11). T2-HighRes approaches outperformed T1 approaches for the detection of early stage atrophy (ES: 0.27 vs 0.10), amyloid positivity (ES: 0.11 vs 0.04), and cognitive associations (ES: 0.22 vs 0.19).

Conclusions: T2-HighRes subfield approaches outperformed whole hippocampus and T1 subfield approaches. None of the different T2-HighRes methods tested had a clear advantage over the other methods. Each has strengths and weaknesses that need to be taken into account when deciding which one to use to get the best results from subfield volumetry.

1. Introduction

Hippocampal neuronal and/or glial dysfunction or neuronal loss severe enough to cause hippocampal volume loss in quantitative magnetic resonance imaging (MRI) is a common feature of many brain disorders, e.g., Alzheimer's disease, multiple sclerosis, epilepsy, schizophrenia, post-traumatic stress syndrome, Parkinson's disease or

traumatic brain injury (Miller and O'Callaghan, 2005; West et al., 1994; Papadopoulos et al., 2009; Bluemcke et al., 1999; Harrison, 2004; McEwen and Magarinos, 1997; Swartz et al., 2006; Pereira et al., 2013). Even non-brain disorders, e.g., hypertension, hypothyroidism or diabetes (Cooke et al., 2014; Petrovitch et al., 2000; Korf et al., 2006) can be associated with hippocampal abnormalities. The hippocampus is not a homogeneous structure but consists of several histologically and

* Corresponding author at: Center for Imaging of Neurodegenerative Diseases, SFGVAMC, Clement Street 4150, San Francisco, CA 94121, USA.

E-mail address: susanne.mueller@ucsf.edu (S.G. Mueller).

¹ Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

functionally specialized but nonetheless tightly interconnected subfields: the subiculum (SUB) which is subdivided into the prosubiculum, subiculum proper, pre- and parasubiculum, the cornu ammonis sectors (CA) 1–3 and the dentate gyrus (DG) (Duvernoy et al., 2013). Animal and histopathological studies have shown that different pathological conditions affect subfields differently, e.g., Alzheimer's disease and hypoxia damage CA1, schizophrenia CA2, traumatic brain injury and post-traumatic stress syndrome CA3 and temporal lobe epilepsy damages the dentate gyrus (e.g., West et al., 1994; Fukutani et al., 2000; Bluemcke et al., 1999; Lucassen et al., 2006; Baldwin et al., 1997; Bluemcke et al., 2007). These observations suggest that subfield specific information might allow for a better differentiation of different disease processes and consequently for an earlier diagnosis than total hippocampal volume (Small, 2014).

The advent of high field (3 Tesla and higher) MRI platforms and the possibility to acquire high resolution images of the hippocampal formation depicting details of its internal structure within a few minutes resulted in the development of a growing number of manual and computational subfield labeling approaches. These techniques were applied to a variety of diseases in small single-site studies and usually found subfield volumetry to be superior to standard whole hippocampal volumetry for the detection of hippocampal damage in the early stages of the disease process, for differentiating between diseases or for the investigation of structure/function relationships (e.g., Wang et al., 2006; Ballmaier et al., 2008; Mueller et al., 2008; Schobel et al., 2009; Neylan et al., 2010; Bender et al., 2013; Kerchner et al., 2014; Schoene-Bake et al., 2014; Chao et al., 2014; Hsu et al., 2015; De Flores et al., 2015; Pluta et al., 2012; Yushkevich et al., 2015b). The promise of hippocampal subfield volumetry techniques however led to two unexpected developments that could potentially limit the usefulness of this approach. One is the lack of standardized boundary definitions for each subfield that complicates the comparison of results from different laboratories. This problem has been recognized by the research community and is now being addressed by an international work group that will develop a harmonized subfield labeling protocol (Yushkevich et al., 2015a; Wisse et al., 2017) using a similar approach as for the harmonization of the outer hippocampus boundaries (Boccardi et al., 2011). The other development is the variety of available labeling techniques (e.g., Pipitone et al., 2014; Kim et al., 2014; Zeineh et al., 2001; Thompson et al., 2004; Chatelat et al., 2008; Goubran et al., 2014). Each of them has its strengths and weaknesses which can complicate the comparison of findings across laboratories but also confuse researchers new to this field who must decide which of the available technique fits their needs best.

The goal of this study was not to compare all existing subfield labeling approaches/protocols but to concentrate on the performance of five commonly used approaches (4 automated, 1 manual labeling) in a common data set. These techniques were selected because each represents a different approach for subfield labeling and because all automated approaches except one are publicly available and used by the research community. The intention was not to determine which of these four approaches captures the disease specific pathology best because this is not possible without a ground truth, i.e. histopathological confirmation, but rather to compare where they find the most prominent differences between groups and how large the differences are. The ideal common data set for such a comparison consists of a large population of well-characterized subjects who show the whole range of normal age-related to severely disease related hippocampal atrophy. This data set became available when the steering committee of the Alzheimer's Disease Neuroimaging Initiative (ADNI) added a high resolution hippocampal sequence to the ADNI MR protocol in a subset of the ADNI sites (cf. Methods).

Based on the image type used for the parcellation two main subfield volumetry approaches can be distinguished. 1. Approaches using a

whole brain T1 weighted image typically at a resolution of 1 mm isotropic or close to that resolution, e.g., shape analysis, radial distance technique, voxel-based morphometry or deformation-based morphometry (Csernansky et al., 1998; Chatelat et al., 2008; Thompson et al., 2004; Yushkevich et al., 2010a). Of the currently available T1-based approaches, the Bayesian inference labeling as implemented in FreeSurfer 5.1. (Van Leemput et al., 2009) and shape analysis based on Large Deformation Diffeomorphic Metric Mapping (Khan et al., 2008) were selected for this project. The former because the algorithm is publicly available and is frequently used, and the second because it was one of the earliest approaches for subfield volumetry that has been continuously refined and optimized for 3 T images. 2. Approaches using a T2 weighted hippocampal image. These images are characterized by a submillimeter in plane resolution but thick slices to increase S/N and therefore depict more details of the internal structure than a T1 image. This has been exploited by manual labeling approaches and increasingly also by automated approaches. Of the currently available approaches ASHS (Yushkevich et al., 2015b) and the latest version of Bayesian inference labeling implemented in FreeSurfer 6.0 (Iglesias et al., 2015) were chosen for this project because both algorithms are publicly available and are used by the larger research community. The manual labeling protocol developed by Mueller et al. (2007) was chosen as manual reference because its limited labeling scheme made it possible to label the over 100 images required for this project within a reasonable amount of time. The selected approaches use different labeling protocols which prevents a direct comparison of the subfield volumes. Instead of focusing on the subset of subfields known to show atrophy and other histopathological features of AD from autopsy studies, e.g., subiculum, CA1, it was therefore decided in a first step to calculate the effect sizes for each and every subfield provided by the approach and compare the performance of the subfields with the largest effect size for each approach regardless of biological plausibility. This provides the type of information that is needed to determine which one of the approaches evaluated here is the most cost efficient in a clinical trial that uses subfield volumetry as an outcome measure. In a second step, it was investigated how well the findings reflected the pattern of subfield atrophy described in histopathological studies of AD. The intention was to compare the performance of these approaches over the whole range of Alzheimer's disease (AD) related hippocampal atrophy and thus the following comparisons were chosen:

1. Ability to detect a group effect for subfield and mesio-temporal volumes in a population consisting of cognitively normal older controls, non-demented subjects with different degrees of mild cognitive impairment and subjects diagnosed with AD (group).
2. Ability to detect subfield and mesio-temporal volumes losses in non-demented subjects with different degrees of mild cognitive impairment (MCI) compared to cognitively normal elderly controls (MCI).
3. Ability to detect an effect of amyloid positivity on subfield and mesio-temporal volumes in cognitively normal subjects (Abeta) which is the stage when treatment with a disease modifying drug would be most efficient.
4. Ability to detect a significant association between general cognitive performance/memory performance and subfield and mesio-temporal volumes in cognitively normal subjects (ADAScog and RVLTL).

2. Methods

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.uci.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to investigate which combination of the measures obtained by serial magnetic resonance imaging (MRI),

positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessments allows for the most efficient monitoring of the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). The initial ADNI was followed by ADNI GO, ADNI 2 and now ADNI 3. To date these 3 projects have recruited > 1500 55–90-year-old participants who were either cognitively normal or diagnosed with early (EMCI), late (LMCI) mild cognitive impairment or with early AD. The ADNI2 hippocampal subfield project was not part of the original ADNI2 project and was made possible by a grant of the Alzheimer's Association. It consisted of three phases. Phase 1: 2011–2012 harmonization and modification (reduction of acquisition time to 8 min) of existing high resolution T2 hippocampus 3 T sequences. Phase 2: January 2012–December 2012: Testing the performance of existing subfield algorithms on ADNI2 imaging data and modification to optimize performance, e.g., development of strategies to allow for partial coverage, testing of different atlases, quality control routines etc. Phase 3. Evaluation phase: January 2013–December 2015. Identification of final study population and data processing

2.1. MRI acquisition

The ADNI2 MRI protocol has been optimized to provide comparable images from different 3 T platforms from the three major vendors Siemens, Philips and General Electric. The high resolution hippocampus sequence was added to the existing ADNI2 imaging protocol of 20 sites with Siemens magnets in December 2012 and acquired at the end of the official protocol. The following images were used for this project: 1. T1 weighted MPRage TR/TE/TI 2300/2.95/900 ms, sagittal, $1.1 \times 1.1 \times 1.2$ mm resolution. 2. T2 weighted turbo spin echo TR/TE 8020/50 ms, inplane resolution 0.4×0.4 mm, slice thickness 2 mm, 28 or 32 slices, coronal, oriented perpendicularly to the long axis of the hippocampus covering the hippocampal head and body but not always the tail in all cases (please see adni.loni.usc.edu/methods/documents/mri-protocols/ Siemens sequences for details).

2.2. Image quality control

684 high resolution exams of 393 different subjects had been acquired between December 2012 and November 2015. The image quality was systematically assessed by trained raters for: hippocampus coverage (full/head and body or incomplete), correctness of FOV angulation regarding the hippocampal axes (meeting specification or not meeting specification), contrast/noise of internal structure of hippocampus (good, moderate, bad), severity of motion artifacts (good, moderate, severe) and other (e.g., artifacts from vessels, pathology). 543 high resolution images from 349 different subjects passed quality control. 84 exams were rejected due to severe motion artifacts which were typically associated with bad contrast/noise. 57 exams were excluded due to positioning errors resulting in incomplete coverage of hippocampal head and body or incorrect angulation. Please see Supplementary material for list of images selected for this project.

2.3. Study population

The images from 106 different subjects were selected for further processing. The reason to focus only on a subset of all available exams was the limited processing capacity of the manual subfield labeling arm that had only one qualified rater. Subjects were selected based on the availability of clinical diagnosis, results of the florbetapir amyloid PET (cut off for amyloid positivity: 1.11) and neuropsychological exam at the timepoint of the selection. Since the intent was to compare the techniques regarding their performance in the early stages of the disease, MCI and healthy controls were preferentially selected. Please see

Table 1
Study population demographics.

Group	Normal	MCI	AD
No	41	57	8
Female/male	22/19	22/35	4/4
Age	75.1 (7.6)	71.1 (7.6)	75.6 (9.0)
Apo E4 pos	10	22	4
SUVr	1.12 (0.21)	1.21 (0.23)	1.41 (0.13)
CDR SB	0.11 (0.38)	1.80 (1.54)	4.56 (1.43)
MMSE	28.8 (1.5)	27.4 (2.8)	21.0 (3.8)
ADAScog	8.4 (4.6)	15.2 (10.0)	31.8 (9.0)
RVLt immediate	47.1 (10.8)	39.7 (13.8)	20.6 (6.1)
RVLt learning	5.8 (2.4)	4.6 (3.1)	2.1 (1.5)
RVLt forgetting	3.8 (3.3)	4.6 (3.1)	4.0 (1.3)

Apo E4 pos, at least one Apo E4 allele; SUVr standardized uptake.

Value ratio relative to cerebellar gray CDR SB, clinical dementia rating scale sum of boxes.

ADAS cog, Alzheimer's Disease Assessment Scale – cognition.

MMSE, Mini-Mental-State-Exam; RVLt, Rey verbal learning test.

Table 1 for demographic information etc. The cognitive functioning of all subjects had been assessed with the standardized ADNI test battery. The Alzheimer's Disease Assessment Scale – cognition (ADAScog) and the immediate recall of the Rey Auditory verbal learning test were chosen from that battery. The former is a measure of global cognitive performance and was expected to show associations with subfields differentiating between cognitively intact and mildly impaired subjects. The latter was chosen as a hippocampus specific measure for encoding/learning and was expected to show an association with CA3 and/or dentate gyrus volume (Mueller et al., 2011). (See Fig. 1.)

2.4. Image processing

2.4.1. High resolution T2 sequences

2.4.1.1. Manual labeling. Input: Resampled high resolution T2 hippocampal image to obtain images where the hippocampal cross-section is perpendicular to the long axis of the left and the right hippocampus.

Output: Left and right label masks in space of resampled image and text file with volumes covered by labels.

Parcellation scheme and atlas: Rview (www.colin-studholme.net/software/software.html) is used for display and labeling. Labels for entorhinal cortex, subiculum, CA1, CA2 and dentate gyrus & CA3 are generated.

Validation: NA.

Customization: Label definition and region to be labeled can be customized.

Quality control: Consistency check, i.e., each label is checked after first pass labeling for labeling accuracy and consistency after processing the complete study population and manually edited if necessary before transferring the volumes into the project database.

Computing power: Standard desktop computer with 2.80GHz CPU and 6.00 GB RAM. The resampling step (preparation and computation of left and right hippocampus) takes ca. 1.5 h/subject, manual labeling ca. 60–90 min/subject for an experienced rater.

Brief method summary: The labeling method including assessment of measurement reliability has been described in detail previously (Mueller et al., 2007, 2010). In brief, the marking scheme depends on anatomical landmarks, particularly on a hypointense line representing myelinated fibers in the stratum moleculare/lacunosum which can be reliably visualized on these high resolution images. Together with external landmarks, e.g., fimbria, collateral sulcus etc. this line is used to identify and manually label the subiculum, the cornu ammonis sectors CA1–CA3 and the dentate gyrus on a length of about 1 cm in the

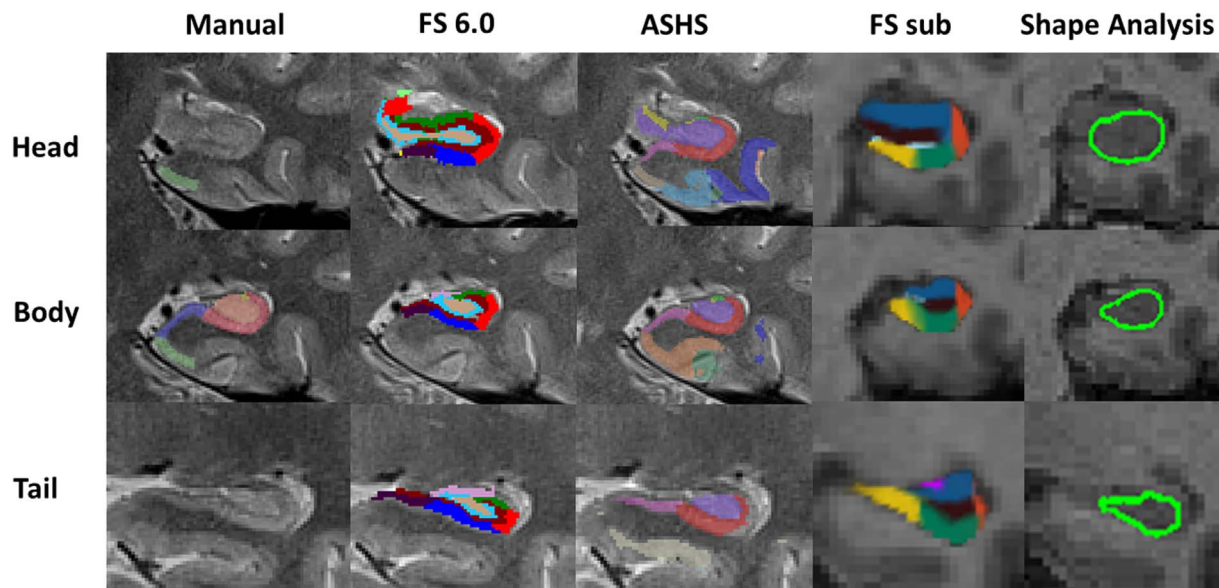


Fig. 1. Parcellation schemes. Manual, manual parcellation labels; green, entorhinal cortex; blue, subiculum; red, CA1; yellow, CA1–2 transition zone; maroon, CA3&dentate. FS 6.0, labels of Freesurfer 6.0 subfield parcellation; yellow, parasubiculum; purple, presubiculum; blue, subiculum; red, CA1, green, CA2/3; tan, CA4; brown, molecular layer; light blue, GC-DG; light green, HATA; lilac, fimbria, ASHS parcellation.

FS Sub, Freesurfer 5.1 subfield parcellation labels; yellow, presubiculum; green, subiculum; red, CA1; blue, CA2/3; bright blue, hippocampal fissure; lilac, fimbria; shape analysis, green, hippocampal boundaries. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

anterior third of hippocampal body. For the sake of efficiency, the labeling section is restricted to the anterior section of the hippocampal body (ca. 60–90 min/subject for an experienced rater).

Limitations: Specially trained rater required, time consuming. Method requires the acquisition of suitable high resolution T2 images.

2.4.1.2. Automated segmentation of hippocampal subfields (ASHS). Input: T1-MRI and high resolution T2 hippocampal image.

Output: Left and right multi-label images and text file with volumes of each subregion.

Parcellation scheme and atlas: Labels for hippocampal subfields CA1–3 and CA4/DG, subiculum, and extrahippocampal cortical regions parahippocampal cortex, entorhinal cortex, BA35 and BA36 together constituting perirhinal cortex, are generated. The atlas provided with the public distribution of the software consists of 28 subjects in whom the subfields were labeled manually by an experienced rater. The atlas population consists of an almost equal proportion of older healthy control and amnesic mild cognitive impairment (aMCI) patients. The method has recently been adapted to 7 T, the latest software release includes a 7 T atlas (Wisse et al., 2016).

Validation: This method has been validated using k-fold cross-validation against manual segmentation in an in-vivo dataset of older healthy controls and aMCI patients (Yushkevich et al., 2015b).

Customization: The method can be adapted regarding number and definition of labels and also regarding atlas images, i.e., it is possible to generate a customized library. Training software is provided as part of the package to utilize such a customized library. A set of 20 atlas images is sufficient.

Quality control: Label images can be displayed overlaid on T2 hippocampal images with ITK snap software (www.itksnap.org) and edited if necessary. Manual editing was not performed on the data presented in this paper.

Computing power: The method is based on a multi-atlas technique that requires multiple pairwise image registrations, which are ideally performed in parallel in a computing cluster. The software provides interface to Sun Grid Engine scheduler to facilitate this. The method

takes 2–3 h to complete on a modern Linux-based cluster.

Brief method summary: This software, named (ASHS) implements a previously published multi-atlas segmentation technique (Yushkevich et al., 2010b). Briefly, the target MRI is registered with a bank of T2-weighted atlas MRIs that includes manually labeled subregions. These manually derived labels were transferred to the target image to produce a set of candidate segmentations. A consensus segmentation is obtained using joint label fusion (Wang et al., 2013a), that takes into account similarity of each atlas image to the target image. Systematic segmentation errors are corrected using a learning-based biased correction technique (Wang et al., 2013b) to generate the final segmentation. Finally, a bootstrapping phase repeats joint label fusion and corrective learning with atlas-to-target registrations seeded by the segmentation result from the previous phase.

Limitations: The atlas provided with the publicly distributed version of the software (<https://www.nitrc.org/projects/ashs/>) is composed of older healthy controls and mild cognitive impairment patients. This atlas is suitable for studies in an Alzheimer's disease population or older healthy subjects, but must be used with caution in a dataset of non-neurodegenerative pathologies and younger subjects, for which the results might be non-optimal, particularly if volumetric analysis is desired. The best results should be achieved with a customized population specific atlas generated from images generated on the same magnet (cf. customization).

2.4.1.3. FreeSurfer 6.0 (FS 6.0). Input: T1-MRI and/or high resolution T2 hippocampal image.

Output: Assuming that both a T1 and a high resolution T2 scan are provided, the algorithm produces a high resolution T2 image in the space of input T1 scan and all other FreeSurfer volumes, a left and right multi-label image with the discrete segmentation of the substructures of each hippocampus, at 0.333 mm resolution, in the physical space of the FreeSurfer T1 data and two text files with the estimated volumes of the hippocampal substructures and of the whole hippocampus. These volumes are computed from probabilistic (rather than binary, winner-take-all) segmentations, so simply counting the number of voxels that

have been labeled as a specific structure in the discrete segmentation will not yield the same result as the volume reported in this file.

Parcellation scheme and atlas: The hippocampal substructures that are segmented by the software are hippocampal tail; parasubiculum; presubiculum; subiculum; CA1; CA2 + CA3; CA4; hippocampus–amygdala transition area (HATA); granule cell layer of dentate gyrus (GC-DG); molecular layer; fimbria; and hippocampal fissure (not included for computing the whole hippocampal volume). These structures were manually labeled in 15 ex vivo scans, and a probabilistic atlas of hippocampal anatomy was built from these delineations, in combination with manual segmentations of 39 1 mm T1 scans of the whole brain at the structure level (i.e., whole hippocampus, whole amygdala, etc.). These 39 scans are important to learn the defining image features of the anatomy of the brain structures surrounding the hippocampus (Iglesias et al., 2015).

Validation: Indirect validation by assessing the ability of the estimated subfield volumes to separate different groups (Iglesias et al., 2015).

Customization: Not possible.

Quality assessment: Labels can be displayed in FreeSurfer's Freeview to assess label accuracy. Editing is theoretically possible but requires advanced knowledge of hippocampal anatomy and was not done for this project.

Computing power: The algorithm takes approximately 30 min to complete, using a single core.

Brief method summary: The method is based on a probabilistic atlas of hippocampal anatomy derived from manual segmentations made on 15 ex vivo scans (including four subjects diagnosed with AD) scanned at (on average) 0.13 mm isotropic resolution. Using a Bayesian algorithm, these manual delineations were then combined with manual segmentations from 39 1 mm T1 scans (including 10 of subjects diagnosed with AD) that had been labeled at the whole structure level, to create a single probabilistic atlas. The atlas is represented as a tetrahedral mesh, in which each node has a corresponding vector of probabilities for the different structures encoded in the atlas. The mesh is adaptive, such that more convoluted regions of the atlas are represented by finer tetrahedra. Once the atlas has been built, segmentation of the hippocampal substructures is posed as a Bayesian inference problem, such that the probability of the segmentation given the input scan and the atlas is maximized. The intensities of the voxels in the scan to analyze are assumed to be samples of a Gaussian mixture model conditioned on the hidden segmentation. The parameters of this model are learned directly from the test scan to segment, which makes the method robust to changes in MRI contrast, and also enables it to immediately generalize to multispectral MRI data (even when some of the channels do not completely cover the hippocampal formation), as described in (Iglesias et al., 2015). Longitudinal labeling has been developed for T1 weighted images (Iglesias et al., 2016) but not been tested for high resolution T2 images with thick slices.

Limitations: The atlas was built using elderly subjects. Its applicability to studies of younger populations or other diseases has not been tested. Finally, a drawback of the ability of this method to adapt to the MRI contrast is that it is not possible to further optimize its performance by training it in a subset of the imaging data to be segmented.

2.4.2. T1 weighted methods

2.4.2.1. Shape analysis. **Input:** T1-MRI and hippocampal label produced by the FreeSurfer recon-all stream.

Output: The primary outputs are binary whole-hippocampus segmentations, corresponding hippocampal 2D surfaces and surface-based subfield delineations. Secondary outputs are multivariate shape indices and summary shape deformation index (univariate).

Parcellation scheme and atlas: The atlas segmentation is a manually traced subset of 74 ADNI1 subjects (not part of investigated cohort).

Validation: The three hippocampal surface zones in the right hemisphere of ten randomly selected subjects were manually outlined and compared with the surface zones as mapped from the template. The intra-class correlation coefficients of the areas of the three surface zones were between 0.90 and 0.97 (Wang et al. 2006).

Customizations: None.

Quality assessment: The method provides a quick-check snapshot generation step for easy quality assurance.

Computing power: Multi-atlas FS-LDDMM requires multiple pairwise image registrations, which are ideally performed in parallel in a computing cluster. On a modern Linux-based cluster, the method takes 2–4 h to complete. A web portal (<http://ceramica.ensc.sfu.ca/>) is provided for collaborators.

Brief method summary: Hippocampal surfaces were generated from T1-weighted images of all subjects using multi-atlas FS-LDDMM (Christensen et al., 2015; Khan et al., 2008; Khan et al., 2013; Wang et al., 2009). ADNI2 EMCI and control subjects that had not been selected for this analysis were used to define EMCI-related surface signature labels. This was done by first computing a population average hippocampal surface and then the vertex-wise deformation for each subject from this average. Vertex-wise generalized linear model (GLM) was used to compare the hippocampal deformation between ADNI2 EMCI and controls. Significant clusters of vertices were identified using Random Field Theory (Taylor and Worsley, 2007; Worsley et al., 1999). The collection of these vertices in each subfield represent “EMCI signature labels.” Next, the hippocampal surfaces of the ADNI2 subjects selected for this project were calculated. The subfield deformation can be represented in a multivariate or a univariate approach. The former uses a PCA to determine the principal components (PC) that account for 80% of the shape variance to represent the surface shape. Each subject's surface is expressed in terms of a linear combination of the PCs with the weights being a multivariate representation of the shape (Csernansky et al., 2004). For this analysis, the univariate approach was used, i.e., the vertex-wise surface deformation corresponding to each EMCI signature label was extracted and the mean calculated. This provided a measure of subfield specific deformation for each subject that is comparable with the other approaches.

2.4.2.2. FreeSurfer subfields version 5.1 (FS sub). **Input:** T1 image.

Output: label image and text file with subfield volumes.

Parcellation scheme atlas: Identifies labels for fimbria, CA2–3, CA1, CA4 and dentate, presubiculum, subiculum, hippocampal fissure, hippocampus (tail) inferior lateral ventricle and choroid plexus based on information from an probabilistic atlas generated from the manual segmentations of the right hippocampus in 10 (6 young, 4 older) cognitively intact subjects (Van Leemput et al., 2009).

Validation: Leave-one-out comparison with manual delineations in ultra-high resolution images.

Customization: Not possible.

Quality control: Each label map was assessed for accuracy, i.e., verified that all subfield labels were within the hippocampus, non-subfield labels, e.g., ventricle label, were not bleeding in the hippocampus and that the hippocampus was completely labeled. No label editing was done, substandard labels were rejected, only labels from images that were rated as pass were used in analysis.

Computing power: 5 h per subject on a 2.83 Hz Intel Xeon E5440 processor.

Brief summary: Automated segmentation of hippocampal subfields using Bayesian inference and a segmentation prior that is defined in the form of a tetrahedral mesh-based probabilistic atlas in which each mesh vertex has an associated vector of probabilities for different hippocampal subfields and surrounding structures of the medial temporal lobe. The subfield routine implemented in FreeSurfer (version 5.1, <https://surfer.nmr.mgh.harvard.edu>) was used for this project.

Limitations: Simplified non-anatomical parcellation scheme (Wisse et al., 2014).

2.4.2.3. FreeSurfer hippocampus version 5.1 (FS Hippo). Input: T1 weighted image.

Output: label mask (aseg_aparc) and text file.

Parcellation and atlas: Subcortical structures including the hippocampi of 39 healthy subjects were manually traced on T1 weighted images. These scans and their segmentations were then used to build a joint probabilistic atlas of labels and intensities.

Customization: Not possible.

Quality control: Each aseg_aparc label map including hippocampal labels was assessed for accuracy. Only hippocampal labels from images that were rated as full pass (all assessed labels met accuracy criteria) were used in this analysis.

Computing power: 5 h per subject on a 2.83 Hz Intel Xeon E5440 processor.

Brief summary: The FreeSurfer processing stream consists of the co-registration of the subject T1 image to an atlas followed by for bias correction, intensity normalization and brain extraction/removal of non-brain tissue. The subcortical structures are labeled by co-registering the subject brain to a probabilistic labeled template in Talairach space. FreeSurfer version 5.1, (<https://surfer.nmr.mgh.harvard.edu>) was used for this project.

Limitations: Only global hippocampal labeling, no subfields.

3. Statistics

All volumes were corrected for intracranial volume (ICV calculated by FreeSurfer) using the following formula: corrected volume = volume \times 1000/ICV. Multiple linear regression analyses with subfield volume as dependent and diagnostic Group, MCI status, SUVR (Abeta) or cognitive measures (ADAScog, RVLt immediate recall) as independent variables of interest and age (all comparisons) and education (cognitive associations only) as independent nuisance variables were performed. The impact of the nuisance variable gender was investigated but found to be non-significant after correcting for ICV and hence not included in the final statistical model. The R^2 from these regression

analyses were used to calculate standardized effect size and power (significance level $\alpha = 0.05$) for the number of the subjects in the actual analysis and not for a fixed number of subjects. Effect sizes are reported rather than the more common p -values because the purpose of this paper was not to report that these subfield approaches are able find significant differences between groups etc. but rather to provide information about how large these differences are (Sullivan and Feinn, 2012). Statistical analyses were done in JMP12 (SAS Institute Inc.) and GPower 3.1 (<http://www.gpower.hhu.de>). SINGLEBAYES.EXE (Crawford and Garthwaite, 2007) was used to identify effect sizes that were significantly higher ($p < 0.05$, one tailed) than others in each comparison.

Considering that all three diagnostic comparisons, i.e., comparisons 1–3, were designed to detect hippocampal atrophy caused by AD, it was expected that the order of the subfields when ranked from largest to smallest effect would be similar for each of these comparisons, e.g., if CA1 had the largest effect size for MCI, it was expected that it would also be the case for Abeta. The consistency of the effect size ranking for the different subfields was investigated using Pearson's correlation coefficients for each side separately, i.e., the effect sizes for comparison 1 were correlated with those from 2 and 3 and those from comparison 2 were correlated with comparison 3 and the mean of the resulting correlation coefficients calculated.

4. Results

4.1. Group comparisons and amyloid effects

Please see Table 2a–2c for detailed effect sizes and power estimates of every subfield. With one exception, the manual labeling approach identified CA1 or CA1-CA2 transition volumes as the subfields with the largest effect sizes for diagnostic and neuropsychological variables of interest. This consistency was also reflected in the high correlation coefficients (mean r left = 0.61, mean r right = 0.61). The FS 6.0 approach identified CA4 and CA1 as having the largest effect sizes for the overall group comparison, fimbria and CA4 as having the largest effect size for the comparison between MCI and cognitively normal and subiculum and molecular layer as having the largest effect size for the

Table 2a
Effect sizes to detect “Group” effects.

a. Effect sizes to detect “Group” effects																			
Method	Side	Measure	PHIPPO	ERC	BA35	BA36	PreSUB	ParaSUB	SUB	CA	CA1	CA1-2	CA3	CA4	Mol Lay	GC-ML-DG	DG	FIMBRIA	HIPPO tail
Manual	L	Power		0.7482					0.9769		1.0000	0.8916					0.9949		
		Effect Size		0.0954					0.1991		0.5122	0.1354					0.2550		
FS 6.0		Power					0.9999	0.4868	1.0000		1.0000		0.9999	1.0000	1.0000	1.0000		0.9959	1.0000
		Effect Size					0.3909	0.0548	0.5131		0.4968		0.3846	0.8700*	0.5090	0.5338		0.2633	0.4165
ASHS		Power	0.9087	0.9497	0.9587	0.9557			0.9594	1.0000							0.9947		
		Effect Size	0.1430	0.1681	0.1762	0.1734			0.1769	0.4306							0.2538		
Shape		Power							0.6993		0.9999						0.9595		
		Effect Size							0.0861		0.5549						0.1769		
FS SUB		Power					1.0000		1.0000				0.9991				0.9998	0.8487	1.0000
		Effect Size					0.4567		0.4567		0.2705		0.3139				0.3644	0.1202	0.5016
FS Hippo		Power																	1.0000
		Effect Size																	0.4857
Manual	R	Power		0.2659					0.9265		0.9999	0.9189					0.9946		
		Effect Size		0.0284					0.1523		0.3784	0.1481					0.2533		
FS 6.0		Power					1.0000	0.6766	1.0000		1.0000		0.9998	1.0000	1.0000	1.0000		0.9360	1.0000
		Effect Size					0.4409	0.0822	0.5675		0.6061		0.3601	0.5545	0.5992	0.5761		0.1582	0.5701
ASHS		Power	0.2292	0.9339	0.8558	0.6910			0.9830	1.0000							0.9998		
		Effect Size	0.0241	0.1568	0.1225	0.0846			0.2110	0.5530							0.3581		
Shape		Power							0.8963		0.8680						0.9599		
		Effect Size							0.1374		0.1265						0.1774		
FS SUB		Power					1.0000		1.0000		0.9517		0.9995				0.9998	0.8294	1.0000
		Effect Size					0.4828		0.5335		0.1698		0.3333				0.3625	0.1146	0.5115
FS Hippo		Power																	1.0000
		Effect Size																	0.5227

Red, subfields with the highest power for $\alpha = 0.05$ and effect size for each method; bold, method with best performance; *, subfield with significantly higher effect size than other subfields within this comparison. PHIPPO, parahippocampus, ERC entorhinal cortex; BA, Brodmann Area; PreSub, presubiculum, ParaSub, parasubiculum, SUB, subiculum, CA, cornu ammonis sector, Mol Lay, molecular layer, GC-ML-DG, granule cell layer of dentate gyrus, DG, dentate gyrus, HIPPO tail, posterior section of hippocampus, Total Hippo, total hippocampal volume from FreeSurfer.

Table 2b

Effect sizes distinction MCI vs. cognitively intact elderly.

b. Effect sizes distinction MCI vs. cognitively intact elders																				
Method	Side	Measure	PHIPPO	ERC	BA35	BA36	PreSUB	ParaSUB	SUB	CA	CA1	CA1-2	CA3	CA4	Mol Lay	GC-ML-DG	DG	FIMBRIA	HIPPO tail	Total Hippo
Manual	L	Power		0.0821					0.0608		0.9329	0.7642					0.8430			
		Effect Size		0.0036					0.0012		0.1553	0.0933						0.1143		
FS 6.0		Power					0.0912	0.0518	0.0568		0.1530		0.7510	0.5674	0.4924	0.6950		0.8821	0.4844	
		Effect Size					0.0046	0.0002	0.0008		0.0112		0.0903	0.0589	0.0489	0.0792		0.1285	0.0479	
ASHS		Power	0.6410	0.2518	0.0500	0.9953			0.2897	0.6543								0.6631		
		Effect Size	0.0699	0.0216	0.5008	0.2697*			0.0256	0.0721								0.0736		
Shape		Power							0.0527		0.2352							0.0742		
		Effect Size							0.0003		0.0198							0.0027		
FS SUB		Power					0.1062		0.2263		0.0519		0.2818				0.3364	0.7752	0.3475	
		Effect Size					0.0062		0.0189		0.0002		0.0248				0.0307	0.0958	0.0319	
FS Hippo		Power																		0.8247
		Effect Size																		0.1087
Manual	R	Power		0.0539					0.0707		0.7093	0.7511					0.9566			
		Effect Size		0.0004					0.0023		0.0819	0.0904						0.1752*		
FS 6.0		Power					0.4681	0.5796	0.4973		0.6197		0.8757	0.9495	0.4964	0.9137		0.7758	0.3754	
		Effect Size					0.0459	0.0606	0.0495		0.0666		0.1259	0.1683*	0.0494	0.1435		0.0959	0.0350	
ASHS		Power	0.0938	0.5445	0.6712	0.2044			0.0553	0.6898								0.9461		
		Effect Size	0.0048	0.0557	0.0750	0.0166			0.0006	0.0783								0.1654*		
Shape		Power							0.8774		0.5762							0.6442		
		Effect Size							0.1266		0.0601							0.0705		
FS SUB		Power					0.5896		0.5256		0.0840		0.5917				0.6580	0.6087	0.4031	
		Effect Size					0.0621		0.0532		0.0038		0.0624				0.0728	0.0649	0.0382	
FS Hippo		Power																		0.8471
		Effect Size																		0.1156

Red, subfields with the highest power for alpha = 0.05 and effect size for each method; bold, method with best performance. *, subfield with significantly higher effect size ($p < 0.05$) than other subfields within this comparison. PHIPPO, parahippocampus, ERC entorhinal cortex; BA, Brodmann Area; PreSub, presubiculum, ParaSub, parasubiculum, SUB, subiculum, CA, cornu ammonis sector, Mol Lay, molecular layer, GC-ML-DG, granule cell layer of dentate gyrus, DG, dentate gyrus, HIPPO tail, posterior section of hippocampus, Total Hippo, total hippocampal volume from FreeSurfer.

Table 2c

Effect sizes for the amyloid effects on hippocampal subfields in cognitively intact amyloid positive and negative controls.

c. Effect sizes for the amyloid effects on hippocampal subfields in cognitively intact amyloid positive and negative controls																				
Method	Side	Measure	PHIPPO	ERC	BA35	BA36	PreSUB	ParaSUB	SUB	CA	CA1	CA1-2	CA3	CA4	Mol Lay	GC-ML-DG	DG	FIMBRIA	HIPPO tail	Total Hippo
Manual	L	Power		0.0596					0.0688		0.1138	0.1185					0.0637			
		Effect Size		0.0034					0.0064		0.0207	0.0222					0.0047			
FS 6.0		Power					0.2297	0.1162	0.3816		0.1451		0.0644	0.0513	0.3301	0.0541		0.0522	0.0718	
		Effect Size					0.0540	0.0215	0.0961*		0.0301		0.0050	0.0005	0.0817	0.0014		0.0008	0.0075	
ASHS		Power	0.3854	0.1453	0.0503	0.0738			0.1827	0.1754							0.0525			
		Effect Size	0.0972*	0.0301	0.0001	0.0081			0.0409	0.0388							0.0009			
Shape		Power							0.0718		0.0990						0.0509			
		Effect Size							0.0051		0.0731*						0.0001			
FS SUB		Power					0.0507		0.0571		0.0527		0.0504				0.0504	0.0571	0.1892	
		Effect Size					0.0002		0.0025		0.0010		0.0001				0.0001	0.0025	0.0427	
FS Hippo		Power																		0.0926
		Effect Size																		0.0142
Manual	R	Power		0.0881					0.1516		0.2842	0.0506					0.2261			
		Effect Size		0.0127					0.0320		0.0690*	0.0002					0.0530			
FS 6.0		Power					0.0954	0.0915	0.0889		0.0907		0.0504	0.0505	0.2157	0.0511		0.0969	0.0702	
		Effect Size					0.0151	0.0138	0.0130		0.0136		0.0002	0.0002	0.0502	0.0004		0.0155	0.0069	
ASHS		Power	0.0644	0.4453	0.0631	0.0521			0.0549	0.2766							0.1136			
		Effect Size	0.0050	0.1145*	0.0045	0.0007			0.0017	0.0669*							0.0207			
Shape		Power							0.1942		0.0688						0.1425			
		Effect Size							0.0324		0.0044						0.0210			
FS SUB		Power					0.0505		0.1061		0.0500		0.1748				0.1007	0.0799	0.0757	
		Effect Size					0.0002		0.0184		0.0000		0.0387				0.0167	0.0101	0.0087	
FS Hippo		Power																		0.0559
		Effect Size																		0.0021

Red, subfields with the highest power for alpha = 0.05 and effect size for each method; bold, method with best performance. *, subfield with significantly higher effect size ($p < 0.05$) than other subfields within this comparison. PHIPPO, parahippocampus, ERC entorhinal cortex; BA, Brodmann Area; PreSub, presubiculum, ParaSub, parasubiculum, SUB, subiculum, CA, cornu ammonis sector, Mol Lay, molecular layer, GC-ML-DG, granule cell layer of dentate gyrus, DG, dentate gyrus, HIPPO tail, posterior section of hippocampus, Total Hippo, total hippocampal volume from FreeSurfer.

comparison between amyloid positive and negative cognitively intact subjects. The inconsistency of the subfields identified as having the largest effect size in these tests was reflected in low correlation coefficients (mean r left = -0.16 , mean r right = -0.16). ASHS identified the CA sector as having the largest effect sizes to detect group effects, BA36 and DG has having the largest effect sizes to detect differences between cognitively intact subjects and subjects diagnosed with MCI and parahippocampal gyrus and entorhinal cortex as having the largest

effect size to detect amyloid associated subfield atrophy. The ranking of the effect sizes and thus consistency of the detected effects was low with $r = -0.27$ on the left side but higher on the right side $r = 0.3$.

T1 based shape analysis identified CA1 and dentate as the subfields with the largest effect sizes for effect of group, and subiculum and CA1 as the subfields with the largest effects sizes to detect differences between cognitively intact and impaired subjects and amyloid positivity. Since the shape analysis only identifies 3 subfields, a formal consistency

Table 3

Effect sizes for the association subfield volumes with cognition in healthy controls: ADAScog and RAVLT immediate.

Effect sizes for the association subfield volumes with cognition in healthy controls: ADAScog and RAVLT immediate																				
Method	Side	Measure	PHIPPO	ERC	BA35	BA36	PreSUB	ParaSUB	SUB	CA	CA1	CA1-2	CA3	CA4	Mol Lay	GC-ML-DG	DG	FIMBRIA	HIPPO tail	Total Hippo
Manual	combined	Power		0.1077					0.4279		0.5980	0.5762					0.1402			
		Effect Size		0.0154					0.0992		0.1531	0.1454					0.0238			
FS 6.0		Power					0.8304	0.1390	0.1776		0.0728		0.0512	0.0839	0.3059	0.1552		0.2671	0.0856	
		Effect Size					0.2671*	0.0234	0.0333		0.0062		0.0003	0.0091	0.0661	0.0276		0.0561	0.0096	
ASHS		Power	0.2602	0.5799	0.4033	0.1795			0.1547	0.4103							0.3579			
		Effect Size	0.0543	0.1467	0.0923	0.0338			0.0274	0.0943							0.0799			
Shape		Power							0.1835		0.0666						0.2711			
		Effect Size							0.0278		0.0036						0.0456			
FS SUB		Power					0.8418		0.3872		0.0519		0.5354				0.6079	0.6137	0.3271	
		Effect Size					0.2756*		0.0878		0.0005		0.1318				0.1567	0.1588	0.0717	
FS Hippo		Power																		0.4395
		Effect Size																		0.1025
Manual	combined	Power		0.0905					0.3383		0.7411	0.7674					0.2327			
		Effect Size		0.0109					0.0746		0.2134	0.2274*					0.0473			
FS 6.0		Power					0.1118	0.0504	0.1035		0.0825		0.2786	0.2479	0.2479	0.5686		0.0937	0.0681	
		Effect Size					0.0164	0.0001	0.0143		0.0087		0.0590	0.0512	0.0512	0.1428		0.0117	0.0049	
ASHS		Power	0.0724	0.1411	0.3114	0.1467			0.0585	0.6512							0.4014			
		Effect Size	0.0060	0.0240	0.0676	0.0254			0.0023	0.1732*							0.0918			
Shape		Power							0.0955		0.2808						0.0538			
		Effect Size							0.0097		0.0476						0.0008			
FS SUB		Power					0.1682		0.3037		0.0565		0.6116				0.6646	0.0924	0.1758	
		Effect Size					0.0309		0.0656		0.0018		0.1580				0.1786*	0.0114	0.0328	
FS Hippo		Power																		0.5810
		Effect Size																		0.1471

Red, subfields with the highest power for $\alpha = 0.05$ and effect sizes for each method; bold, method with best performance. *, subfield with significantly higher effect size ($p < 0.05$) than other subfields within this comparison. PHIPPO, parahippocampus, ERC entorhinal cortex; BA, Brodmann Area; PreSub, presubiculum, ParaSub, parasubiculum, SUB, subiculum, CA, cornu ammonis sector, Mol Lay, molecular layer, GC-ML-DG, granule cell layer of dentate gyrus, DG, dentate gyrus, HIPPO tail, posterior section of hippocampus, Total Hippo, total hippocampal volume from FreeSurfer. ADAScog, Alzheimer's Disease Assessment Scale – cognition; RAVLT, immediate recall of the Rey Auditory verbal learning test.

assessment with Pearson correlation coefficients was not possible. FS Sub identified the fimbria as having the largest effect sizes for the differentiation between cognitively intact and impaired subjects, subiculum and hippocampal tail as having the largest effect sizes to detect a group effect and hippocampal tail and CA2/CA3 as having the largest effect size to detect amyloid effects in cognitively intact controls. The ranking of the effect sizes was low on the left side with $r = -0.003$ and moderate on the right side with $r = 0.26$. With exception of the effect size for group effect, the effect sizes of total hippocampal volume derived from FS Hippo was lower than those seen by the T2 high resolution image based approaches. The means of the subfields with the highest power and effect sizes for each T2 method (values in red in Table 2a–2c, Table 3) were generally higher than those of the T1 based methods for all three comparisons (cf. Table 4).

A well acknowledged problem of all subfield labeling approaches including those chosen for this project is the lack of a common set of rules to identify their boundaries which makes a direct comparison of the findings across different labs difficult (Yushkevich et al. 2015). Although the identification of a common set of rules for subfield labeling was beyond the scope of this project, it is possible to combine labels from more detailed protocols to mimic the parcellations of less detailed protocols. The

Table 4

Comparison of power and effect sizes of T2 and T1 based subfield approaches.

	Mean power			Mean effect size			Mean power	Mean effect size
	High Res T2	Whole brain T1	P-value	High res T2	Whole brain T1	P-value		
Group	1.0000	0.9900	0.24	0.5584	0.4419	0.33	1.0000	0.5042
MCI	0.9438	0.6241	0.02	0.1771	0.0768	0.01	0.8359	0.1122
Abeta	0.3051	0.1643	0.06	0.0749	0.0467	0.18	0.0742	0.0081
Cog	0.6659	0.5146	0.26	0.1850	0.1369	0.37	0.5102	0.1248

High Res T2 includes manual labeling, ASHS and Freesurfer 6.0 subfields.

Whole Brain T1 includes shape analysis and Freesurfer 5.1 subfields.

P-value, p for comparison T2 vs T1 subfield approach.

less detailed protocol used by ASHS and shape analysis consisted of a CA label, a DG label and a subiculum label. The CA label was mimicked for the manual approach by combining the CA1 and CA1–2 transition zone volumes, for FS 6 by combining the volumes for subiculum, CA1, CA2/3 and molecular layer and for FS Sub by combining the volumes for subiculum, CA1, CA2/3. The presubiculum label of FS 6.0 and FS Sub corresponds to the subiculum label of ASHS, shape analysis and the manual approach. The DG label for FS 6.0 was mimicked by combining the labels of CA4 and granular cell layer dentate. The DG estimates for the manual labeling was based on its CA3&DG volume and those for FS sub on its CA4-DG label. Fig. 2 shows the effect sizes for these compounded labels. Using the compounded labels, all three high resolution T2 based approaches identified left and right CAC as the subfield with the highest effect size for group. Manual labeling and FS 6.0 identified left and right CAC and ASHS right CAC but left subiculum with left CAC as a close second as the most affected combined label for Abeta effects. All three identified either CAC or DG as the most affected subfield for the MCI effects. Taken together, simplifying parcellation schemes by compounding labels increased the consistency within each approach and across the three approaches. FS SUB identified left and right subiculum as the region with the largest group effect and left and right DG as the region with the largest MCI effect. Right CAC and left subiculum with left CAC as a close second showed the largest effect sizes for Abeta with the FS SUB approach. The use of the compounded labels also increased the consistency of the findings within FS SUB but not of those of high resolution T2 based approaches.

4.2. Subfield volumes as predictor of cognitive function in cognitively intact subjects

Left and right volumes were averaged for this assessment. Please see Table 3 for the detailed results. The manual approach identified CA1 and CA1–2 transition as having the largest effect sizes for ADAScog and RAVLT immediate recall. FS 6.0 identified the presubiculum as the subfield with the largest effect for ADAScog and the dentate granular cell layer volume as the subfield with the largest effect size for immediate RAVLT recall. ASHS found that ERC had the largest effect size for

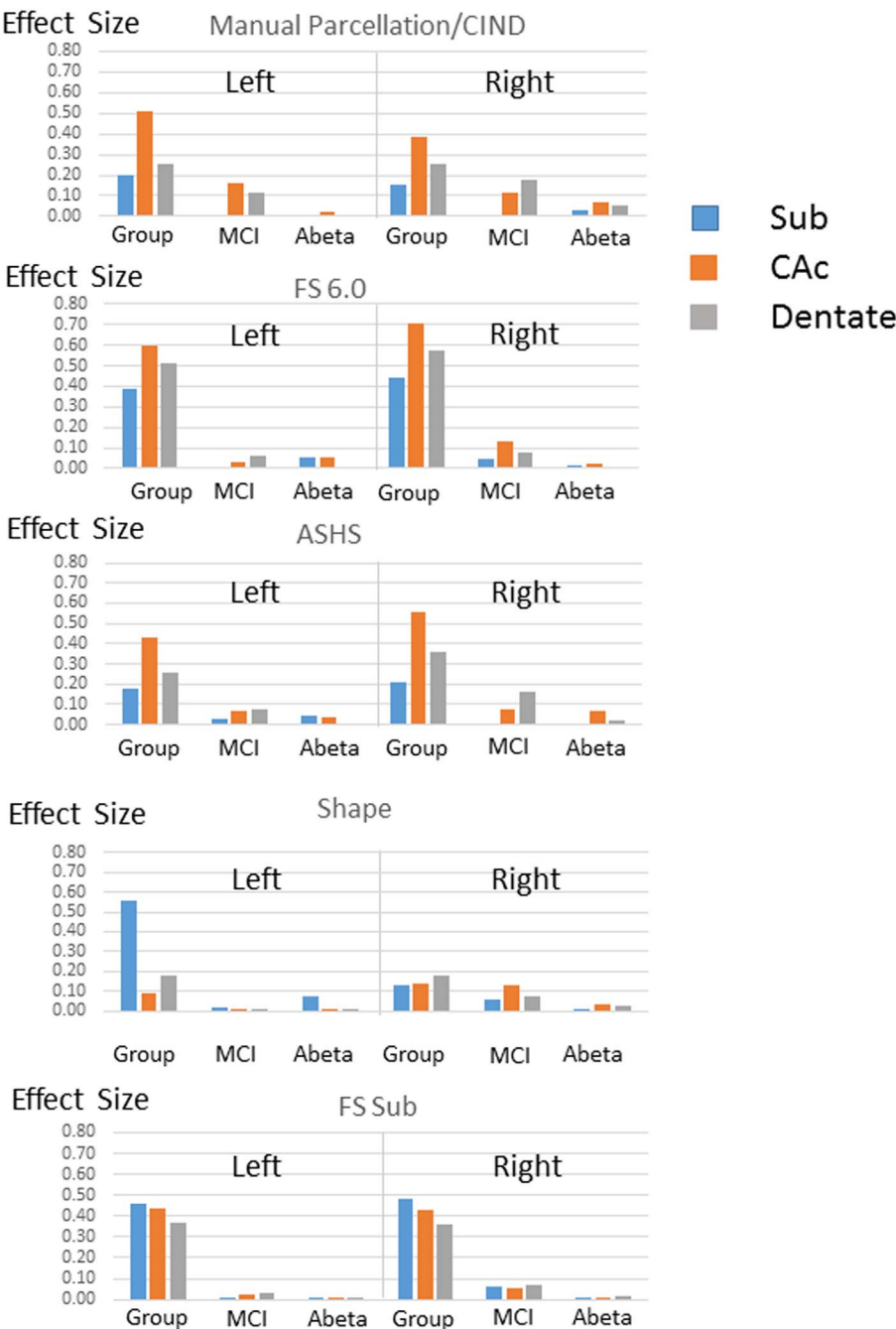


Fig. 2. Bar plots of effect sizes for combined labels. CA corresponds to all cornu ammonis labels, DG corresponds to the dentate gyrus and Sub to the subiculum. Please see text for a description how these labels were generated for each of the approaches. “Group” refers to the ability to detect a group effect in a population consisting of cognitively normal elderly controls, non-demented subjects with different degrees of mild cognitive impairment and subjects diagnosed with Alzheimer’s Disease. ‘MCI’ refers to the ability to detect a group effect on subfield and mesio-temporal volumes in a population consisting of cognitively normal elderly controls, and non-demented subjects with different degrees of mild cognitive impairment. ‘Abeta’ is the ability to detect an effect of amyloid positivity on subfield and mesio-temporal volumes in cognitively normal subjects.

ADAScog and that CA had the largest effect size for immediate RVLt. Shape analysis identified the dentate gyrus as having the largest effect size for ADAScog and CA1 as the subfield with the largest effect size for RVLt immediate recall. Finally, FS Sub subfields showed the same pattern as FS 6.0, i.e., presubiculum had the highest effect size for ADAScog and CA4&dentate had the highest effect size for RVLt immediate recall. The means of the subfields with the highest power and effect sizes for each T2 method (values in red in Table 2a–2c, Table 3) were generally higher than those of the T1 based methods for all both comparisons (cf. Table 4).

5. Discussion

The comparison revealed two major findings. 1. It was possible to acquire T2 high resolution hippocampal images in a large multi-site project that were comparable regarding image quality and accurate and

consistent positioning to images acquired for dedicated small research projects where a one-to-one training and immediate feedback regarding quality are possible. This is an important step for this technique's successful transition from a research tool to a clinical routine procedure. 2. Using effect size as criterion subfield volumetry but especially approaches using a T2 high resolution image outperformed traditional hippocampal volumetry regarding their ability to detect conditions characterized by subtle hippocampal atrophy, i.e., comparisons between MCI vs CN, amyloid neg vs pos, and associations subfield cognition/memory in cognitively intact subjects. T2 high resolution approaches also outperformed subfield approaches using the whole brain T1 image for the comparisons between MCI vs CN and amyloid neg vs pos. However, the T2 high-resolution approaches identified different subfields as having the highest effect sizes for these comparisons. This was not unexpected given the differences of the labeling scheme. Reducing the differences of parcellation schemes between the three T2

high resolution approaches by combining labels and generating compounded labels eliminated left/right differences within each method and increased the consistency for the identification of the most affected region across methods. None of the different T2 high resolution methods tested had a clear advantage over the other methods. Each has strengths and weaknesses that need to be considered when deciding which one to use to get the best results from subfield volumetry in AD.

Attempts to use high resolution T2 images to measure subfield volumes have been made since the introduction of the hippocampal unfolding technique by Zeineh et al. (2001) and have been intensified after 3 T magnets became widely available in academic hospitals and research institutions in the last decade. Despite its popularity for research and the development of reliable, fast and computationally efficient automated subfield volumetry approaches quantitative hippocampal subfield volumetry has yet to make the transition from research to clinical application. One of the reasons for the delay is the relatively lengthy acquisition time of about 9–13 min which is rather long for a busy clinical setting. For this project, this issue was addressed by shortening the acquisition time to about 8 min. Another reason is the need for careful positioning that requires some knowledge of hippocampal anatomy. Despite detailed instructions incorrect positioning with insufficient coverage of the hippocampus was a problem in the early phase of this project and resulted in the exclusion of about 8% of all acquired images. The problem was eliminated after adding the positioning information as a tab to the information displayed at the MR console. The images acquired for ADNI2 were of similar quality as those acquired in research settings. 12% were rated as not suitable for processing mostly due to motion artifacts or poor contrast in the hippocampal formation which compares well with the percentage of images excluded from processing in this population in research settings. Taken together, it can be concluded that it is possible to obtain high resolution T2 hippocampal images of suitable quality for manual and automated subfield labeling within a reasonable acquisition time in a multi-site project which makes it possible to use this sequence in imaging protocols for drug trials or for clinical purposes.

There are two ways to assess the performance of different subfield volumetry approaches. The first is statistical efficiency which compares approaches based on effect size or minimal number of subjects needed to find any statistical difference. The second is biological plausibility which addresses the question to what degree the measurement with the highest effect size corresponds to the known histopathological features of the disease process. AD is characterized by a well-defined progression of its histopathological hallmarks amyloid plaques and neurofibrillary tangles. Particularly the latter has been shown to be closely associated with cognitive performance and neuron loss within the hippocampus and mesio-temporal lobe (Giannakopoulos et al., 1996; Giannakopoulos et al., 2007; Fukutani et al., 2000). The earliest site of neurofibrillary tangle accumulation in the preclinical state of AD is the trans-entorhinal cortex (BA 35) followed by the entorhinal cortex. The tangles then progress to the hippocampus where they cause the most prominent neuron loss in CA1 followed by that in the subiculum while the dentate gyrus is relatively spared in the early clinical stages (West et al., 1994). Subfield volumetry should ideally show the same kind of progression, i.e., the most prominent volume loss in the trans-entorhinal cortex/entorhinal cortex in the early preclinical phase followed by volume loss in CA1 (MCI and early AD) and subiculum (early AD) and affect the dentate/hilus region in the more advanced stages. Based on these considerations, CA1 and entorhinal cortex would be expected to show the highest effect sizes in all comparisons except for “group” where subiculum and eventually dentate gyrus are also affected.

Regarding statistical efficiency the findings of this project replicated those of previous studies (e.g., De Flores et al., 2015; Mueller et al., 2010) that showed that subfield volumetry based on a T2 high

resolution image outperforms hippocampus volumetry for the detection of subtle and more localized atrophy characterizing the early stages of AD as evidenced by the power and effect size calculations for “MCI” and “Abeta”. This was expected given that the increased accuracy for the detection of subtle atrophy was one of the major reasons for the development of subfield volumetry. Nonetheless, it is satisfying that this assumption was confirmed for all T2 high resolution approaches investigated here.

Approaches based on a high resolution T2 hippocampal image had typically higher effect sizes than subfield approaches using the lower resolution whole brain T1. This finding was also not unexpected. The myelinated tissue of the stratum moleculare and lacunosum appears as a very characteristic hypointense line in this image and its acquisition parameters are usually set to maximize this contrast in the population of interest. In contrast, T1 based subfield approaches use standard whole brain T1 weighted images with a lower resolution whose contrast has not been optimized for the hippocampus. On such images the stratum moleculare and lacunare is slightly hyperintense compared to the surrounding tissue and far less prominent although it is possible to enhance it by acquiring and averaging multiple T1 images. Without any strong information from the internal structure though, subfield parcellation by FS SUB is mostly guided by the outer hippocampal boundaries and by the information from the probabilistic prior. This limitation and the simplified non-anatomical parcellation scheme reduce the ability of FS SUB to detect subtle disease related atrophy and explain why the subfields with highest effect sizes for the different comparisons, do not reflect the known atrophy pattern very well despite large effect sizes. The rationale behind subfield volumetry based on hippocampal shape analysis inherently acknowledges the limited information about the internal structure in the T1 image and tries instead to optimize the information provided by the outer boundaries. However, the ability of surface-based approaches to detect subtle disease related effects is on the one hand limited by the existence of physiological shape variants, e.g., the varying number of digitations in the hippocampal head or rotational variants of the body, and on the other hand by the fact that abnormalities restricted to a specific subfield are often not well reflected on the surface. In addition, the shape measure used in this analysis was the univariate summary index, selected to be comparable to other methods. Multivariate approaches are better suited to capture shape differences due to group differences. These limitations explain why shape analysis had the lowest effect sizes of the subfield specific approaches for comparisons assessing subtle atrophy. It is important to emphasize in this context that the area where shape analysis is likely to outperform the other approaches, i.e. its ability to identify the region of maximal atrophy along the anterior-posterior axis, was not tested in this project.

If effect size is used as the sole criterion, none of the different high resolution hippocampal T2 based approaches consistently outperformed the other approaches, i.e., the effect size differences between the best performing labels of the three T2 based approaches were small and mostly not significant. This leaves the question how well they were able to detect the atrophy pattern associated with the disease process. One of the major criticisms of the manual labeling approach used for this project is that the labeling is restricted to the anterior third of the hippocampal body. While there is no doubt that this approach would miss localized atrophy in the posterior body, its effect sizes for the detection of a less focal atrophy pattern compared well with the automated whole hippocampus labeling approaches. The manual labeling identified CA1 or the neighboring CA1–2 transition area as the most prominently affected subfields except for the “MCI” comparison in the right hippocampus where the largest effect size was found in the dentate gyrus and CA3 area. The ranking of the effect sizes for the manual labeling was also quite consistent across the comparisons. It is possible

that the limited labeling approach that excluded difficult to trace regions such as the hippocampal head and tail contributed to the consistency of the findings. ERC however had often the smallest effect sizes in these comparisons. The poor performance of ERC for the manual labeling was not unexpected. The re-slicing of the image that is necessary to optimize the cross-section for hippocampal labeling and the limited coverage of the ERC often aggravate partial volume artifacts in the ERC region and thus the non-disease related variability. This was different for ASHS. Except for the “group” comparison that identified CA as the most affected subfield and the “cognitively normal vs MCI” comparison that identified the right dentate gyrus as having the largest effect size, ASHS identified ERC, neighboring Brodmann Area (BA) 36 or the parahippocampus as the regions with the highest effect size with CA usually a close second. This is not surprising since ASHS is the only labeling approach evaluated here that has been optimized for entorhinal/perirhinal volumetry. It is the only approach that has a separate label for the transentorhinal cortex or BA 35 (Taylor and Probst, 2008) although the other, larger entorhinal/perirhinal labels outperformed the BA 35 volume in all comparisons. A downside of the detailed entorhinal/perirhinal labels was that it affected the consistency of the ranking order of the effect sizes. FS 6.0 uses the most detailed parcellation for hippocampal subfield labeling of all approaches investigated here. The detailed labeling makes the assessment of this approach somewhat challenging because there exist no histopathological studies that systematically investigated the distribution of AD related abnormalities across layers of different subfields. As seen in the assessment of the other high resolution T2 based methods, the subfields showing the largest effect sizes varied between comparisons and the consistency of the effect size order across comparisons was low. The higher number of labels in FS 6.0 likely contributed to the low consistency. Reducing the complexity of the labeling scheme by combining labels increased the consistency of the findings between left and right hippocampal volumes for each T2 approach and also across T2 approaches but reduced the effect sizes.

The observations regarding detailed labeling schemes in the previous paragraph point to another important requirement for biologically meaningful subfield labeling. The input images need to consistently depict a minimal set of landmarks for these parcellation schemes to work well. This minimal number is higher for more detailed parcellations than for less sophisticated schemes. Furthermore, outputs from automated subfield labeling routines should undergo a rigorous visual quality control by raters who know the hippocampal anatomy well enough to spot labeling inaccuracies. Such a quality control step becomes more complicated with elaborate parcellation schemes that require expert knowledge. These concerns raise the question if these detailed rhinal and hippocampal parcellations of ASHS and FS 6.0 should not be reserved for high quality 3 T images with motion correction or even 7 T images that depict the landmarks reliably and are addressing hypotheses that require such a detailed parcellation while projects that do not acquire this kind of high quality image and do not have layer specific hypotheses should consider reducing the complexity of the labeling by combining to enhance reliability and reproducibility across projects.

As mentioned before, each of T2 high resolution methods has its set of strengths and weaknesses. Manual segmentation is time consuming and therefore not an option for the processing of large datasets encompassing hundreds of images. It also requires experienced, specially trained rater(s) who are not available everywhere. However, this weakness can also turn into a strength because experienced raters can adapt more easily to physiological and pathological shape and contrast variants of the hippocampus than atlas-based automated subfield labeling approaches. ASHS, on the other hand, is based on multi-atlas segmentation, which has become a very popular segmentation

approach in recent years, and produces state-of-the-art results in many medical image segmentation tasks. However, ASHS relies on the accuracy of an intensity-matched registration of the images to be segmented to a library of labeled images which is likely to suffer if the former has been acquired with a different sequence or platform than the atlas images. In the case of ASHS, this can be overcome by generating population and magnet specific libraries, but this usually requires again the input of expert raters who select the appropriate images and edit the library labels. FS 6.0's main advantage is that its intensity models are learned directly from the individual scan to be segmented, which makes it less dependent on magnet type and sequence. However, FS 6.0 uses the probabilistic information from an ex vivo atlas for this process and thus its performance will be impaired in cases where the intensity distribution deviates too much from the atlas due to shape variants or due to the disease processes.

In conclusion, subfield volumetry outperformed hippocampal volumetry in their ability to detect subtle atrophy that characterizes the early stages or preclinical stages of AD: T2 high resolution based approaches performed better than T1 based approaches but the latter have the advantage that they do not require the acquisition of a dedicated image. The automated T2 based subfield labeling approaches tested in this project compared well with the manual approach and their accuracy is likely to further improve in the future. However, just as for manual labeling approaches, the accuracy of the automated algorithms depends on the quality of the image that is being labeled and on how well the atlas or a priori information used by the algorithm reflects the process of interest.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.nicl.2017.12.036>.

Acknowledgment

The project was supported by a grant of the Alzheimer's Association (ADNI 2-12-233036) to MWW and SGM. None of the authors has a conflict of interest to declare.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering (U01 AG024904), and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd. and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

- Baldwin, S.A., Gibson, T., Callihan, C.T., Sullivan, P.G., Palmer, E., Scheff, S.W., 1997. Neuronal cell loss in the CA3 subfield of the hippocampus following cortical contusion utilizing optical disector method for cell counting. *J. Neurotrauma* 14, 385–398.
- Ballmaier, M., Narr, K.L., Toga, A.W., Elderkin-Thompson, V., Thompson, P.M., Hamilton, L., Haroon, E., Pham, D., Heinz, A., Kumar, A., 2008. Hippocampal morphology and distinguishing late-onset from early onset elderly depression. *Am. J. Psychiatry* 165, 229–237.
- Bender, A.R., Daugherty, A.M., Raz, N., 2013. Vascular risk moderates associations between hippocampal subfield volumes and memory. *J. Cogn. Neurosci.* 25, 1851–1862.
- Bluemcke, I., Beck, H., Lie, A.A., Wiestler, O.D., 1999. Molecular neuropathology of human mesial temporal lobe epilepsy. *Epilepsy Res.* 36, 205–223.
- Bluemcke, I., Pauli, E., Clusman, H., Schramm, J., Becker, A., Elger, C., Merschhemke, M., Meenke, H.J., Lehmann, T., von Deimling, A., Scheiwe, C., Zentner, J., Volk, B., Romstoeck, J., Stefan, H., Hildebrandt, M., 2007. A new clinico-pathological classification system for mesial temporal lobe sclerosis. *Acta Neuropathol.* 113, 235–244.
- Boccardi, M., Ganzola, R., Bocchetta, M., Pievani, M., Redolfi, A., Bartzokis, G., Camicioli, R., Csernansky, J.G., de Leon, M.J., de Toledo-Morrell, L., Killiany, R.J., Lehericy, S., Pantel, J., Pruessner, J.C., Soininen, H., Watson, C., Duchesne, S., Jack, C.R., Frisone, G.B., 2011. Survey of protocols for the manual segmentation of the hippocampus: preparatory steps towards a joint EADC-ADNI harmonized protocol. *J. Alzheimers Dis.* 26 (Suppl. 3), 61–75.
- Chao, L.L., Kriger, S., Buckley, S., Ng, P., Mueller, S.G., 2014. Effects of low-level sarin and cyclosarin exposure on hippocampal subfields in Gulf War veterans. *Neurotoxicology* 44, 263–269.
- Chatelat, G., Fouquet, M., Kalpouzos, G., Denghien, I., De la Sayette, V., Viader, F., Mezenge, F., Landeau, B., Baron, J.C., Eustache, F., Desgranges, B., 2008. Three dimensional surface mapping of hippocampal atrophy progression from MCI to AD and over normal aging as assessed using voxel-based morphometry. *Neuropsychologia* 46, 1721–1731.
- Christensen, A., Alpert, K., Rogalski, E., Cobia, D., Rao, J., Beg, M.F., Weintraub, S., Mesulam, M.M., Wang, L., 2015. Hippocampal subfield surface deformity in non-semantic primary progressive aphasia. *Alzheimers Dement.* 1, 14–23.
- Cooke, G.E., Mullally, S., Correia, N., O'Mara, S.M., Gibney, J., 2014. Hippocampal volume is decreased in adults with hypothyroidism. *Thyroid* 24, 433–440.
- Crawford, J.R., Garthwaite, P.H., 2007. Comparison of a single case to a control or normative sample in neuropsychology: development of a Bayesian approach. *Cognitive Cogn. Neuropsychol.* 24, 343–372.
- Csernansky, J.G., Joshi, S., Wang, L., Haller, J.W., Gado, M., Miller, J.P., Grenander, U., Miller, M.I., 1998. Hippocampal morphometry in schizophrenia by high dimensional brain mapping. *Proc. Natl. Acad. Sci.* 95, 11406–11411.
- Csernansky, J.G., Wang, L., Joshi, S.C., Ratnanather, J.T., Miller, M.I., 2004. Computational anatomy and neuropsychiatric disease: probabilistic assessment of variation and statistical inference of group difference, hemispheric asymmetry, and time-dependent change. *NeuroImage* 23 (Suppl. 1), S56–68 (2004).
- De Flores, R., La Joie, R., Chételat, G., 2015. Structural imaging of hippocampal subfields in healthy aging and Alzheimer's disease. *Neuroscience* 309, 29–50.
- Duvernoy, H.M., Catlin, F., Risold, P.Y., 2013. *The Human Hippocampus*, 4th Edition. Functional Anatomy, Vascularization and Serial Sections with MRI Springer, Berlin, Heidelberg, New York.
- Fukutani, Y., Cairns, N.J., Shiozawa, M., Sasaki, K., Sudo, S., Isaki, K., Lantos, P.L., 2000. Neuronal loss and neurofibrillary degeneration in the hippocampal cortex in late-onset sporadic Alzheimer's disease. *Psychiatry Clin. Neurosci.* 54, 523–529.
- Giannakopoulos, P., Hof, P.R., Koevari, E., Vallet, P.G., Herrmann, F.R., Bouras, C., 1996. Distinct pattern of neuronal loss and Alzheimer's disease lesion distribution in elderly individuals older than 90 years. *J. Neuropathol. Exp. Neurol.* 55, 1210–1220.
- Giannakopoulos, P., von Gunten, A., Kövari, E., Gold, G., Hermann, F.R., Hof, P.R., Bouras, C., 2007. Stereological analysis of neuropil threads in the hippocampal formation: relationship with Alzheimer's disease neuronal pathology and cognition. *Neuropathol. Appl. Neurobiol.* 33, 334–343.
- Goubran, M., Rudko, D.A., Santyr, B., Gati, J., Szekeres, T., Peters, T.M., Khan, A.R., 2014. In vivo normative atlas of the hippocampal subfields using multi-echo susceptibility imaging at 7 Tesla. *Hum. Brain Mapp.* 35 (8) (3588–601).
- Harrison, P.J., 2004. The hippocampus in schizophrenia: a review of the neuropathological evidence and its pathophysiological implications. *Psychopharmacology* 174, 151–162.
- Hsu, P.J., Shou, H., Benzinger, T., Marcus, D., Durbin, T., Morris, J.C., Sheline, Y., 2015. Amyloid burden in cognitively normal elderly is associated with preferential hippocampal subfield volume loss. *J. Alzheimers Dis.* 45, 27–33.
- Iglesias, J.E., Augustinack, J.C., Ngyuen, K., Player, C.M., Player, A., Wright, M., Roy, N., Frosch, M.P., McKee, A., Wald, L.L., Fischl, B., van Leemput, K., 2015. A computational atlas of the hippocampal formation using ex vivo ultra-high resolution MRI. Application to adaptive segmentation of in vivo MRI. *NeuroImage* 115, 117–137.
- Iglesias, J.E., Van Leemput, K., Augustinack, J., Insausti, R., Fischl, B., Reuter, M., Alzheimer's Disease Neuroimaging Initiative, 2016. Bayesian longitudinal segmentation of hippocampal substructures in brain MRI using subject-specific atlases. *NeuroImage* 141, 542–555.
- Kerchner, G.A., Berdnik, D., Shen, J.C., Bernstein, J.D., Fenesy, M.C., Deutsch, G.K., Wyss-Coray, T., Rutt, B.K., 2014. Apo E4 worsens hippocampal CA1 apical neuropil atrophy and episodic memory. *Neurology* 82, 691–697.
- Khan, A.R., Wang, L., Beg, M.F., 2008. FreeSurfer-initiated fully-automated subcortical brain segmentation in MRI using large deformation diffeomorphic metric mapping. *NeuroImage* 41, 735–746.
- Khan, A.R., Wang, L., Beg, M.F., 2013. Multistructure large deformation diffeomorphic brain registration. *IEEE Trans. Biomed. Eng.* 60, 544–553.
- Kim, H., Bernhardt, B.C., Kulaga-Yoskovitz, J., Caldarou, B., Bernasconi, A., Bernasconi, N., 2014. Multivariate hippocampal subfield analysis of local MRI intensity and volume: application to temporal lobe epilepsy. *Med. Image Comput. Comput. Assist. Interv.* 17, 170–178 (17170–8).
- Korf, E.S., White, L.R., Scheltens, P., Launer, L.J., 2006. Brain aging in very old man with type 2 diabetes: the Honolulu Asia Aging Study. *Diabetes Care* 10, 2268–2274.
- Lucassen, P.J., Heine, V.M., Muller, M.B., van der Beek, E.M., Wiegant, V.M., De Kloet, E.R., Joels, M., Fuchs, E., Swaab, D.F., Czeh, B., 2006. Stress, depression and hippocampal apoptosis. *CNS Neurol. Disord. Drug Targets* 5, 531–546.
- McEwen, B.S., Magarinos, A.M., 1997. Stress effects on morphology and function of the hippocampus. *Ann. N. Y. Acad. Sci.* 821, 271–284.
- Miller, D.B., O'Callaghan, J.P., 2005. Aging, stress and the hippocampus. *Ageing Res. Rev.* 4, 123–140.
- Mueller, S.G., Stables, L., AT, Du, Schuff, N., Truran, D., Cashdollar, N., Weiner, M.W., 2007. Measurement of hippocampal subfields and age-related changes with high resolution MRI at 4T. *Neurobiol. Aging* 28, 719–726.
- Mueller, S.G., Schuff, N., Rapoport, S., Elman, J., Weiner, M.W., 2008. Selective effect of ApoE4 on CA3 and dentate in normal aging and Alzheimer's disease using high resolution MRI at 4T. *NeuroImage* 42, 42–48.
- Mueller, S.G., Schuff, N., Yaffe, K., Madison, C., Miller, B., Weiner, M.W., 2010. Hippocampal atrophy pattern in mild cognitive impairment and Alzheimer's disease. *Hum. Brain Mapp.* 31, 1339–1347.
- Mueller, S.G., Chao, L.L., Berman, B., Weiner, M.W., 2011. Evidence for functional specialization of hippocampal subfields detected by MR subfield volumetry on high resolution images at 4T. *NeuroImage* 56, 851–857.
- Neylan, T.C., Mueller, S.G., Wang, Z., Metzler, T.J., Lenoci, M., Truran, D., Marmar, C.R., Weiner, M.W., Schuff, N., 2010. Insomnia severity is associated with a decreased volume of CA3/dentate gyrus hippocampus. *Biol. Psychiatry* 68, 494–496.
- Papadopoulos, D., Dukes, S., Patel, R., Nicholas, R., Vora, A., Reynolds, R., 2009. Substantial archaocortical atrophy and neuronal loss in multiple sclerosis. *Brain Pathol.* 19, 238–253.
- Pereira, J.B., Junque, C., Bartes-Faz, D., Ramirez-Ruiz, B., Marti, M.J., Tolosa, E., 2013. Regional variability of hippocampal subfields and memory deficits in Parkinson's disease. *Hippocampus* 23, 720–728.
- Petrovitch, H., White, L.R., Izmirlian, G., Ross, G.W., Havlik, R.J., Markesberry, W., Nelson, J., Davies, D.G., Hardman, J., Foley, D.J., Launer, L.J., 2000. Midlife blood pressure and neuritic plaques, neurofibrillary tangles and brain weight at death: the HAAS Honolulu-Asia Aging Study. *Neurobiol. Aging* 21, 57–62.
- Pipitone, J., Park, M.T., Winterburn, J., Lett, T.A., Lerch, J.P., Pruessner, J.C., Lepage, M., Voineskos, A.N., Chakravarty, M.M., 2014. Multi-atlas segmentation of the whole hippocampus and subfields using multiple automatically generated templates. *NeuroImage* 101 (494–51).
- Pluta, J., Yushkevich, P., Das, S., Wolk, D., 2012. In vivo analysis of hippocampal subfield atrophy in mild cognitive impairment via semi-automatic segmentation of T2-weighted MRI. *J. Alzheimers Dis.* 31, 85–99.
- Schobel, S.A., Lewandowski, N.M., Corcoran, C.M., Moore, H., Brown, T., Malaspina, D., Small, S.A., 2009. Differential targeting of the CA1 subfield of the hippocampal formation by schizophrenia and related psychotic disorders. *Arch. Gen. Psychiatry* 66, 938–946.
- Schoene-Bake, J.C., Keller, S.S., Niehuismann, P., Volmering, E., Elger, C., Deppe, M., Weber, B., 2014. In vivo mapping of hippocampal subfields in mesial temporal lobe epilepsy: relation to histopathology. *Hum. Brain Mapp.* 35, 4718–4728.
- Small, S.A., 2014. Isolating pathogenic mechanism embedded within the hippocampal circuit. *Neuron* 84, 32–33.
- Sullivan, G.M., Feinn, R., 2012. Using effect size –or why the p value is not enough. *J. Grad. Med. Educ.* 3, 279–282.
- Swartz, B.E., Houser, C.R., Tomiyasu, U., Walsh, G.O., DeSalles, A., Rich, J.R., Delgado-Escueta, A., 2006. Hippocampal cell loss in posttraumatic human epilepsy. *Epilepsia* 47, 1373–1382.
- Taylor, K.I., Probst, A., 2008. Anatomic localization of the transentorhinal region of the perirhinal cortex. *Neurobiol. Aging* 29, 1591–1596.
- Taylor, J.E., Worsley, K.J., 2007. Detecting sparse signals in random fields, with an application to brain mapping. *J. Am. Stat. Assoc.* 102, 913–928.
- Thompson, P.M., Hayashi, K.M., De Zubicaray, G.L., Janke, A.L., Rose, S.E., Semple, J., Hong, M.S., Herman, D.H., Gravano, D., Doddrell, D.M., Toga, A.W., 2004. Mapping hippocampus and ventricular change in Alzheimer disease. *NeuroImage* 22, 1754–1766.
- Van Leemput, K., Bakker, A., Benner, T., Wiggins, G., Wald, L.L., Augustinack, J., Dickerson, B.C., Golland, P., Fischl, B., 2009. Model-based segmentation of hippocampal subfields from ultra-high resolution in vivo MRI. *Hippocampus* 19, 549–557.
- Wang, L., Miller, J.P., Gado, M.H., McKeel, D.W., Rithermich, M., Miller, M.I., Morris, J.C., Csernansky, J.G., 2006. Abnormalities of hippocampal surface structure in very mild dementia of the Alzheimer's type. *NeuroImage* 30, 52–60.
- Wang, L., Khan, A., Csernansky, J.G., Fischl, B., Miller, M.I., Morris, J.C., Beg, M.F., 2009. Fully-automated, multi-stage hippocampus mapping in very mild Alzheimer disease. *Hippocampus* 19, 542–548.
- Wang, H., Suh, J.W., Das, S.R., Pluta, J.B., Craige, C., Yushkevich, P.A., 2013a. Multi-atlas segmentation with joint label fusion. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 611–623.
- Wang, H., Das, S.R., Pluta, J., Craige, C., Altinay, M., Avants, B.B., Weiner, M., Mueller, S.G., Yushkevich, P.A., 2013b. A learning-based wrapper method to correct systematic errors in automatic image segmentation: consistently improved performance in hippocampus, cortex and brain segmentation. *NeuroImage* 55 (968–85), 2013.
- West, M.J., Coleman, P.D., Flood, D.G., Troncoso, J.C., 1994. Differences in the pattern of hippocampal neuronal loss in normal ageing and Alzheimer's disease. *Lancet* 344, 769–772.
- Wisse, L.E., Biessels, G.J., Geerlings, M.I., 2014. A critical appraisal of the hippocampal subfield segmentation package in FreeSurfer. *Front. Aging Neurosci.* 6 (261). <http://dx.doi.org/10.3389/fnagi.2014.00261>. (Sept 25).
- Wisse, L.E., Kuijff, H.J., Honingh, A.M., Wang, H., Pluta, J.B., Das, S.R., Wolk, D.A.,

- Zwanenburg, J.J., Yushkevich, P.A., Geerlings, M.I., 2016. Automated hippocampal subfield segmentation at 7T MRI. *AJNR Am. J. Neuroradiol.* 37, 1050–1057.
- Wisse, L.E., Daugherty, A.M., Olsen, R.K., Berron, D., Carr, V.A., Stark, C.E., Amaral, R.S., Amunts, K., Augustinack, J.C., Bender, A.R., Bernstein, J.D., Brocardi, M., Bocchetta, M., Burggren, A., Chakravarty, M.M., Ekstrom, A., de Flores, R., Insausti, R., Kanel, P., Kedo, O., Kennedy, K.M., Kercher, G.A., LaRocque, K.F., Liu, X., Maass, A., Malykhin, N., Mueller, S.G., Ofen, N., Palombo, D.J., Parekh, M.B., Pluta, J.B., Pruessner, J., Raz, N., Rodrigue, K.M., Schoemaker, D., Shafer, A.T., Steve, T.A., Suthana, N., Wang, L., Winterburn, J.L., Yassa, M.A., Yushkevich, P.A., La Joie, R., Hippocampus Subfield Group, 2017. A harmonized protocol for hippocampal and parahippocampal subregions: why do we need one and what are the key goals? *Hippocampus* 27, 3–11.
- Worsley, K.J., Andermann, M., Koulis, T., MacDonald, D., Evans, A.C., 1999. Detecting changes in nonisotropic images. *Hum. Brain Mapp.* 8, 98–101.
- Yushkevich, P.A., Avants, B.B., Das, S.R., Pluta, J., Altinay, M., Craige, C., 2010a. Alzheimer's Disease Neuroimaging Initiative. In: Bias in estimation of hippocampal atrophy using deformation-based morphometry arises from asymmetric global normalization: an illustration in ADNI 3 T MRI data. *Neuroimage* 50, pp. 434–445.
- Yushkevich, P.A., Wang, H., Pluta, J., Das, S.R., Craige, C., Avants, B.B., Weiner, M.W., Mueller, S., 2010b. Nearly automated segmentation of hippocampal subfields in vivo focal T2 weighted MRI. *NeuroImage* 53 (1028–1024).
- Yushkevich, P.A., Amaral, R.S., Augustinack, J.C., Bender, A.R., Bernstein, J.D., Boccardi, M., Bocchetta, M., Burggren, A.C., Carr, V.A., Chakravarty, M.M., Chetelat, G., Daugherty, A.M., Davachi, L., Ding, S.L., Ekstrom, A., Geerlings, M.I., Hassam, A., Huang, Y., Iglesias, J.E., La Joie, R., Kercher, G.A., LaRocque, K.F., Libby, L.A., Malykhin, N., Mueller, S.G., Olsen, R.K., Palombo, D.J., Parekh, M.B., Pluta, J.B., Preston, A.R., Pruessner, J.C., Raganath, C., Raz, N., Schlichting, M.L., Schoemaker, D., Singh, S., Stark, C.E., Suthana, N., Tompay, A., Turowski, M.M., van Leemput, K., Wagner, A.D., Wang, L., Winterburn, J.L., Wisse, L., Yassa, M.A., Zeineh, M.M., 2015a. Hippocampal Subfields Group (HSG). Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in vivo MRI: towards a harmonized segmentation protocol. *NeuroImage* 111, 526–541.
- Yushkevich, P.A., Pluta, J.B., Wang, H., Xie, L., Ding, S.L., Gertje, E.C., Manusco, L., Kilot, D., Das, S.R., Wolk, D.A., 2015b. Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. *Hum. Brain Mapp.* 36, 258–287.
- Zeineh, M.M., Engel, S.A., Thompson, P.A., Bookheimer, S.Y., 2001. Unfolding the human hippocampus with high resolution and structural and functional MRI. *Anat. Rec.* 265, 111–120.